International Journal of Computer Science Engineering
and Information Technology Research (IJCSEITR)
ISSN(P): 2249-6831; ISSN(E): 2249-7943
Vol. 4, Issue 4, Aug 2014, 1-8

ANALYZING LARGE-SCALED APPLICATIONS IN CLOUD COMPUTING ENVIRONMENT

DEEPTI SINGH, DEEPTI SARASWAT & N S RAGHAVA

Department of Information Technology, Delhi Technological University, Delhi, India

ABSTRACT

Cloud computing has snatched a large amount of consideration from all over the globe, because of its highly demanding features. It has evolved as a new platform, providing computing as a utility service to all its users. We are particularly concerned about applications built using the cloud computing paradigm, where data is maintained and processed in multiple datacenters. These datacenters are located in geographically distant regions. Some popular examples of such applications include Facebook, Yahoo, Google Apps, etc. Workload distribution in such large-scaled applications are very important from the view of the system's performance. Geographical location of the datacenters and the user groups provides a major framework for Cloud Service Providers (CSPs) to distribute load among different servers in datacenters. In this paper, we have discussed some important issues which must be considered while developing large-scaled application using cloud technology. This paper also gives an insight upon three load balancing policies in combination with various broker policies, so as to choose the best combination according to the choice of different Cloud Service Providers in case of large-scaled applications. Overall analysis has been done using CloudAnalyst [5].

KEYWORDS: Cloud Analyst, Cloud Computing, Load Balancer, Load Balancing, Service Broker

INTRODUCTION

One of the fastest growing areas in computing research is cloud computing. It has given a new idea of utilizing the power of computing like other utility services of gas and electricity [1]. Cloud technology offers services that are classified into three major groups: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). These services are used as per the users demand and Cloud Service Provider decides as how to avail these services to different users. Large-scaled applications on internet like Facebook, Yahoo, Google, e-commerce applications can achieve maximum benefits from cloud technology. Every Cloud Service Provider (CSP) aims, providing high quality services to end users with minimum data processing and resource utilization cost and maximum resource utilization. With the advent of the cloud computing it is now possible to achieve a better trade-off between all these parameters, but a matter of concern is that how these large-scaled applications be configured over the Internet. There are a number of issues and different performance metrics [8] which need great concern of developers and researchers in deploying any such large-scaled applications over the internet. Based on pay-per-use model, large scaled computing infrastructure can be provided on rent to users, at a very low price. This feature helps in reducing the cost which is required during initial-setup of any application. Cloud technology allows its users to use its services elastically, which facilitates users to scale up or down these services according to their will. These large-scaled application systems can take great advantage of such cloud services. The main aim of such applications is to minimize setup cost and to deliver a high service quality to the end users [2].

Issues Effecting the Performance of Large-Scaled Applications

There are a number of issues which must be addressed carefully so as to achieve a better performance in cloud computing systems. A cloud service provider should have an idea about these issues and their effect. Major issues which must be considered in a cloud computing environment to achieve a better performance of the overall system are discussed below:

Geographical Location

The geographical location of datacenters highly affects the overall performance of the large-scaled application system. Geographical analyses help in choosing the suitable place for datacenter deployment. Cloud Analyst [5] divides the world into six major areas that overlap approximately with the six main continents of the world. Failure from single region can be protected by other regions where the datacenters have been created for the same purpose.

VM (Virtual Machine) Load Balancing Policy

We have considered three load balancing techniques, namely Round Robin, Throttled and Equal Spread Current Execution for simulation purpose. All of these policies have some pros and cons. They are chosen by different cloud service providers, according to the requirement of the system. These techniques are discussed below:

Round Robin (RR) Policy: As the name suggests Round Robin Policy, handles the requests of the clients in a circular manner on a first come first serve basis. This algorithm is quite simple in terms of complexity in comparison to the other two algorithms discussed below.

Equal Spread Current Execution (ESCE) Policy [5]: This policy equally spreads the current execution load among all the available virtual machines equally. The Load Balancer maintains a record of all virtual machines and the number of allocations assigned to each virtual machine. When a new request comes, Load Balancer searches the least loaded VM.

Throttled Policy [5]: Throttled policy, assigns each VM only one job at one time and other jobs are assigned only when the current job has completed successfully. The job manager entity maintains a list of all virtual machines. Allocation of virtual machine to the appropriate job is done by using the indexed list. A job is assigned to the machine only if the job is appropriate for that machine, else job manager entity delays the user request and takes the job in waiting queue for fast processing.

Service Broker Policy

A Service Broker entity in Cloud Analyst [5] routes the user requests coming from different user groups located at different geographical locations to datacenters in the cloud. The service broker policy can be of following types [3]:

Closest Datacenter Policy: This Policy, selects the closest datacenter in terms of Network Latency. Network latency is taken from the latency matrix held in the Internet Characteristics in Cloud Analyst [5]. For simulation purpose Amazon EC2 [7] has been taken as a standard for all network characteristics.

Optimize Response Time Policy: In this policy, response time is the major criteria. All the datacenters are actively monitored by the service broker entity. Then the broker sends the request to the datacenter, which gives best response time to the end user at the time it is queried.

Dynamic Configuration Policy: In this policy, the service broker is assigned an extra task of scaling the application deployment depending upon the load it currently had. This policy scale-up or scale-down virtual machines dynamically in the datacenters according to the best processing time. If the current processing time is less than best processing time, then best processing time is updated consequently.

Peak Hours Analysis

Network traffic in each region is different, so it is necessary to analyze peak working hours in each region. Also network traffic is high during peak hours. In Cloud Analyst [5], we can put region-wise peak hours to make the simulation process more real. We have done simulation based upon the statistics of peak hours in each region [4].

Simulation Configuration & Experimental Setup

All the experimental work has been done using Cloud Analyst. Cloud Analyst [5] is a GUI based simulator for modeling and analysis of large scaled applications. It is fabricated on top of Cloud Sim toolkit, by outspreading Cloud Sim functionality. Figure.1 shows the basic component upon which Cloud Analyst [5] has been built.

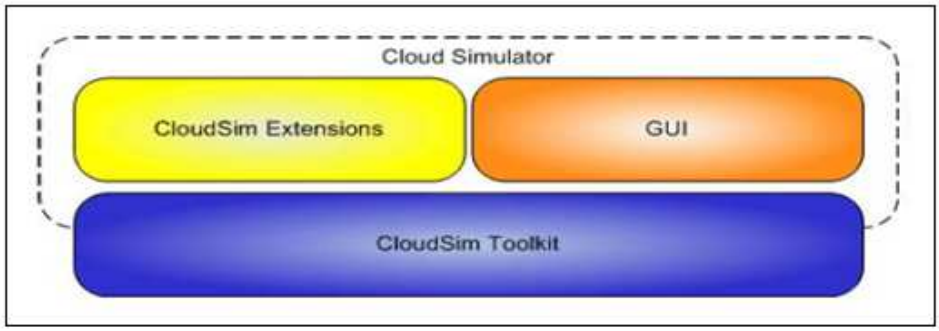

Figure 1: Cloud Analyst Fabricated on Top of CloudSim Toolkit

Through the GUI, its different components like Userbase, Internet, Service Broker Policy, Internet Cloudlet, Datacenter Controller and VM Load Balancing Policies can be configured differently for different cloud users. One of the major social-networking giant- Facebook, is having millions of users over the globe. We have used the Facebook user statistics for simulating load balancing policies in combination with different service broker policies. Registered users on Facebook on 31-03-2012 are listed in Table 1. [6]

Table 1: Registered Users on Face Book as on 31-03-2012

Major Geographic Region	Registered Users	CloudAnalyst Region Id
North America	173,284,940	R0
South America	112,531,100	R1
Europe	232,835,740	R2
Asia	195,034,380	R3
Africa	40,205,580	R4
Oceania/Australia	13,597,380	R5

We have considered a similar system, but at $1/10^{th}$ of the scale of the Facebook, for our simulation task. We have created six user bases defining the above mentioned six major geographic regions. Table 2 represents these user bases with some important parameters used in the simulation.

Table 2: User Bases with Parameters

User base	Region Id	Time Zone	Peak Hours (GMT)	Simultaneous Online Users during Peak Hours	Simultaneous Online Users during off-Peak Hours
UB1	R2	GMT + 1.00	20:00 – 22:00	400,000	40,000
UB2	R3	GMT + 6.00	01:00 – 03:00	300,000	30,000
UB3	R0	GMT - 6.00	13:00 – 15:00	150,000	15,000
UB4	R1	GMT – 4.00	15:00 – 17:00	100,000	10,000
UB5	R4	GMT + 2.00	21:00 – 23:00	50,000	5,000
UB6	R5	GMT + 10.00	09:00 – 11:00	40,000	4,000

In order to provide the simulation framework more real environment, assumptions regarding the network configurations and the pricing of resources were made similar to one of the most popular cloud services provider Amazon EC2 [7]. According to the user statistics we have assumed different number of requests from each user in each hour and also different data size request per user so as to achieve results closer to real environment. We have used certain parameters which are fixed during entire simulation. These parameters are listed in Table 3.

Table 3: Fixed Parameters Used during Simulation

Parameters	Value Used
Simulation duration	10 Hours
Data centers	2
User Bases	6
Number of Hosts (Each having 4 processors)	20
Virtual Machine	50 for each Datacenter
User grouping factor in user bases	1000
Request grouping factor in datacenters	100
Executable instruction length per request	250 bytes

RESULTS

Results are analyzed in six different cases based upon the configuration stated in the previous section. Three load balancing policies namely Round Robin (RR), Equal Spread Current Execution (ESCE) and Throttled are combined one by one with two service broker policies namely Closest Datacenter and Optimize Response Time. All the six cases are discussed below:

Case-1: Round Robin Policy with Closest Datacenter Policy

In this case load balancing policy used is Round Robin and service broker policy is Closest Datacenter, a simulated output is shown in Figure 2. Region-wise response time is generated for each user base. In this configuration, total cost of data center one (DC1) comes out to be 1124.30 $ much higher than the cost of datacenter two (DC2), which is only 267.23 $.

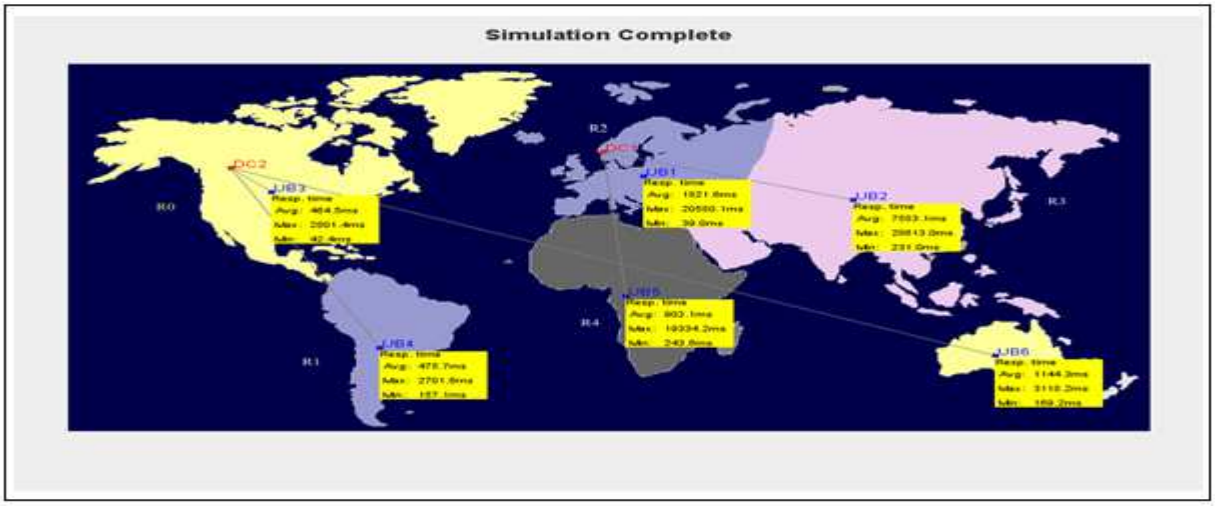

Figure 2: Simulation Completed for Round Robin Policy with Closest Datacenter Policy

Case-2: Round Robin Policy with Optimize Response Time Policy

In this case we change our service broker policy to Optimize Response Time and load balancing to the application is done through Round Robin policy. Figure 3 shows the simulated output for this configuration. Fine lines in the Figure 3 shows several requests from different user bases to the specific datacenter. As compared to previous case overall response time in this situation is higher than the previous case, but the total cost for both the datacenter is balanced.

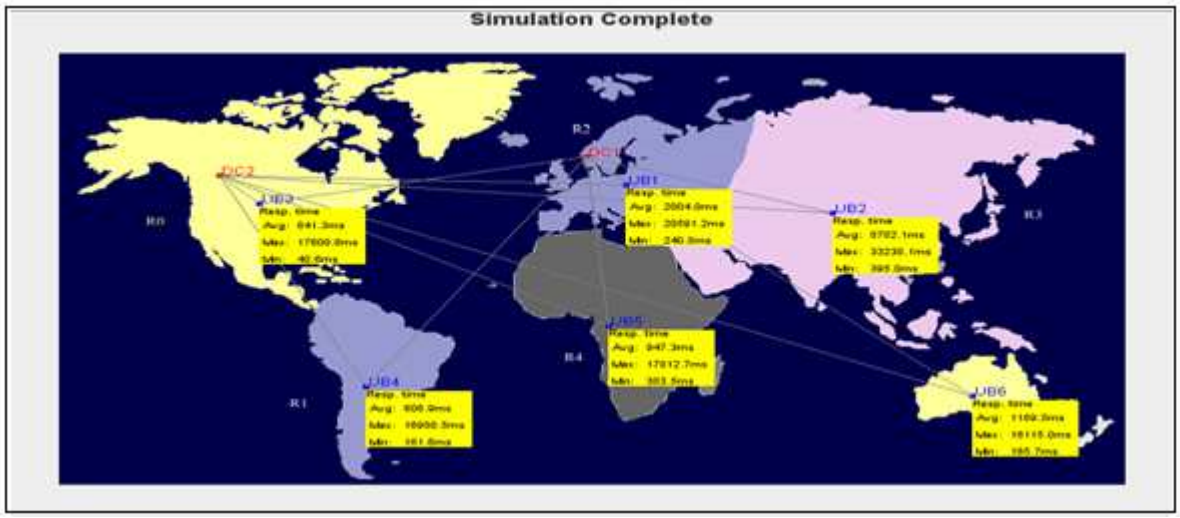

Figure 3: Simulation Completed for Round Robin Policy with Optimized Response Time Policy

CASE-3: Equal Spread Current Execution Policy with Closest Datacenter Policy

Here we changed the load balancing policy to ESCE and service broker policy is Closest Datacenter. The methodologies of all these policies have been discussed earlier. In this case data processing time and overall response time is better than the previous cases and the total cost is similar to CASE-1. In Figure 4 the whole situation for this configuration can be analyzed clearly.

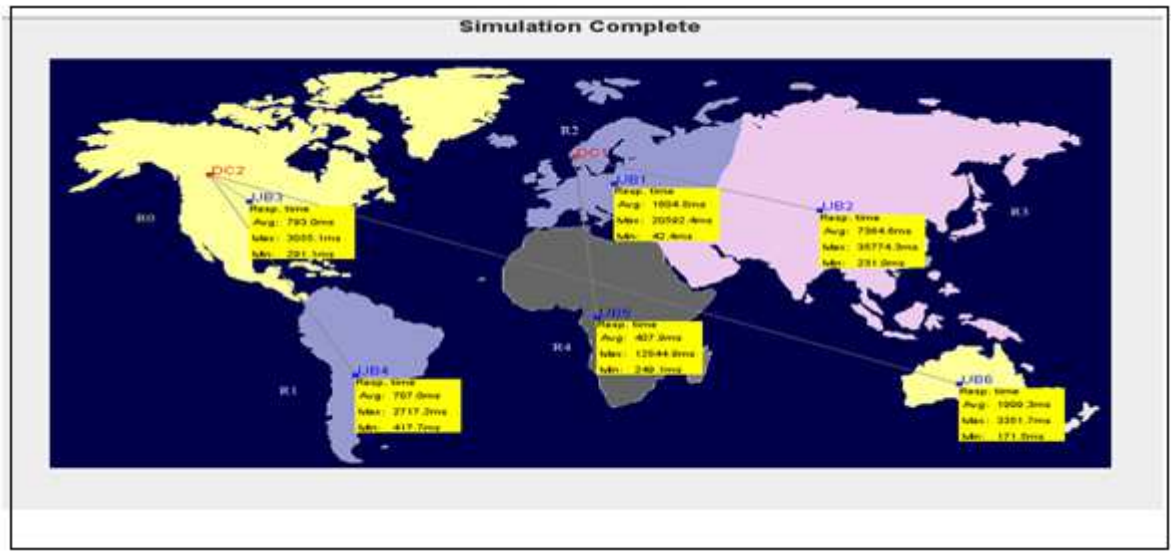

Figure 4: Simulation Completed for Equal Spread Current Execution Policy with Closest Datacenter Policy

CASE-4: Equal Spread Current Execution Policy with Optimize Response Time Policy

In this case, the load balancing policy is same as in CASE-3, but service broker policy is changed to Optimize Response Time. Overall response time and data processing is much higher than all the cases. Also the cost for all virtual machines and data transfer is high. Therefore, we can say that this configuration will not provide any benefit to end users and CSPs. Figure 5 shows the simulated output for this case.

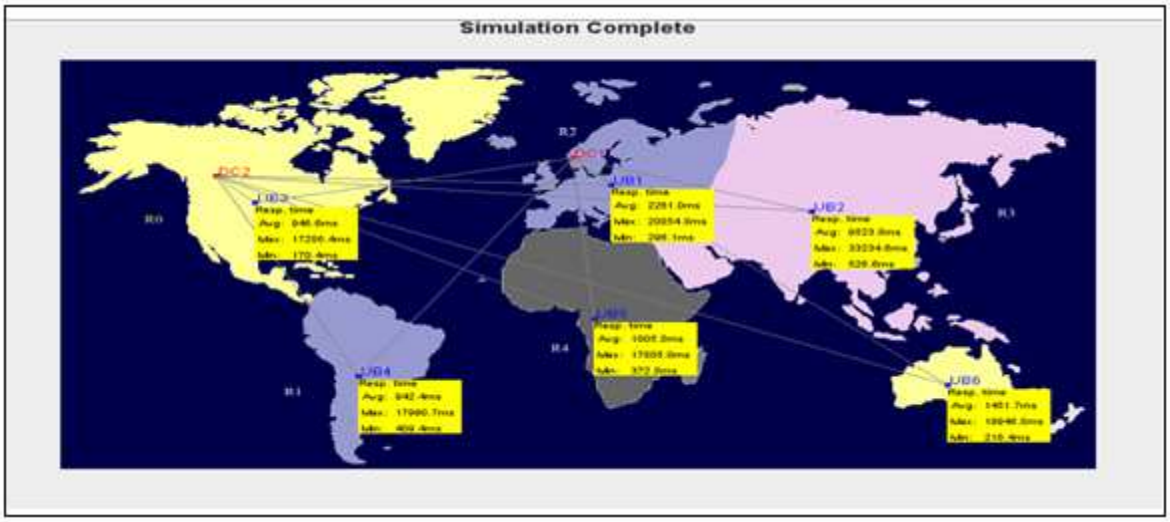

Figure 5: Simulation Completed for Equal Spread Current Execution Policy with Optimize Response Time Policy

CASE-5: Throttled Policy with Closest Datacenter Policy

Here the load balancing is done through Throttled policy and service broker policy is Closest Datacenter. This configuration gives the much better result than the previous cases. Figure 6 shows the response time for different user bases. There is a major downfall in response time as well as in data processing time at each datacenter. Total cost is same for all the cases where service broker policy used is Closest Datacenter.

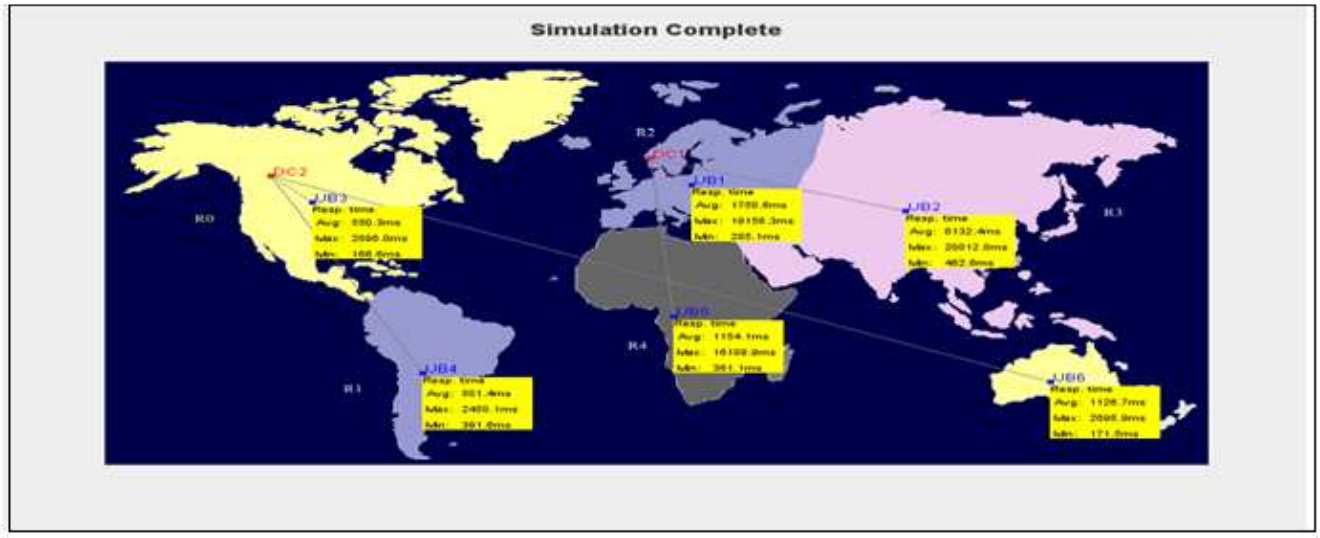

Figure 6: Simulation Completed for Throttled Policy with Closest Datacenter Policy

CASE-6: Throttled Policy with Optimize Response Time Policy

In this case, the load balancing policy is Throttled, but service broker policy has changed to Optimize Response Time. Overall response time and the data processing time are 3034.47 ms and 2760.83 ms respectively, which is lower as compared to all the previous cases. And the total cost for all virtual machines and data transfer is also minimized. Therefore, we can say this configuration is best for our application. It will benefit end-users with minimum response time and also profit cloud service providers with minimum cost of operation. Figure 7, shows the simulated result in this case.

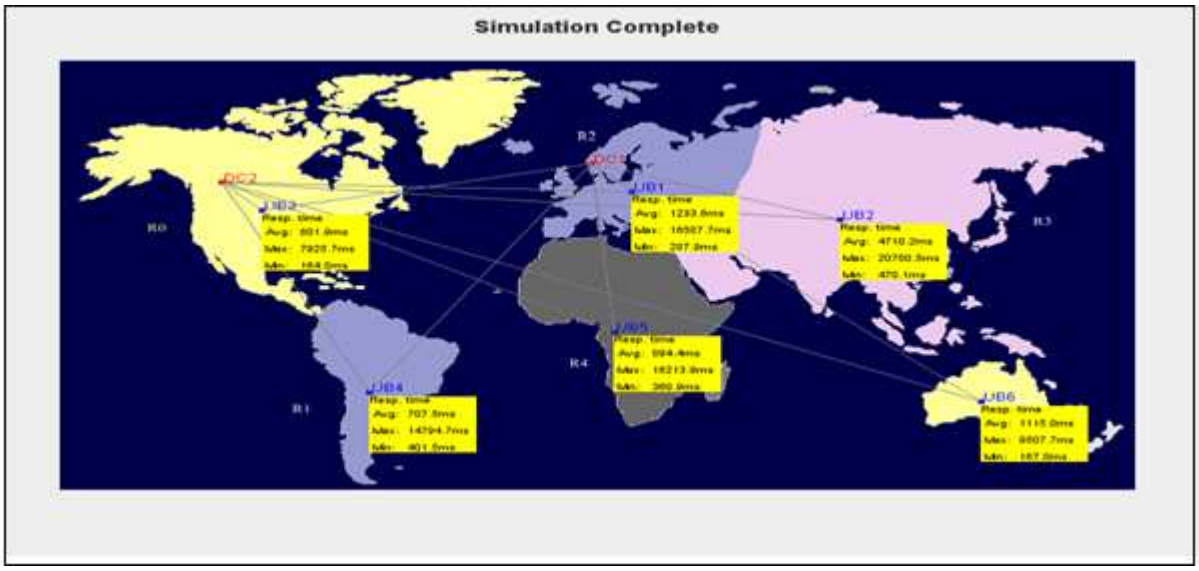

Figure 7: Simulation Completed for Throttled Policy with Optimize Response Time Policy

The overall response time in each case can be analyzed graphically through Figure 8. The minimum, maximum and average response time in each case is shown separately for each case. Generally, we consider the average response time for evaluating the performance of the overall system.

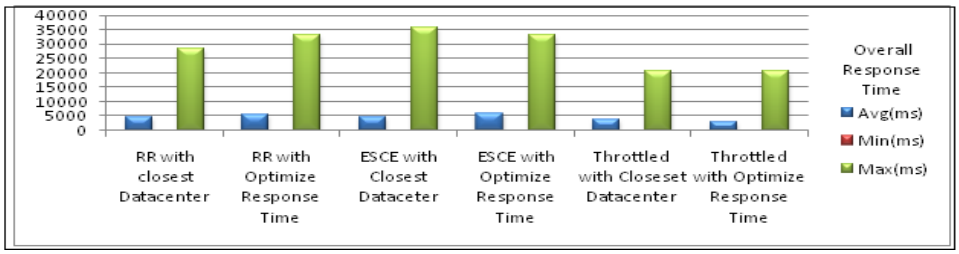

Figure 8: Overall Response Time for all the Cases

It can be clearly observed from the results that Throttled load balancing policy is best suited for our application. It gives the minimum response time compared to others. And for service broker policy, we can say Closest Datacenter policy gives the maximum profit to the cloud service provider by giving the lowest cost for all virtual machines and data transfer.

CONCLUSIONS

It is really difficult to simulate large-scaled applications over the internet. There are various parameters upon which the performance of the application is tested and validated. Right choice of load balancing and service broker policies can provide major improvements in the performance of these large-scaled applications. A Cloud Service Provider wants to utilize maximum resources with minimum cost and end-user wants their task to be done in minimum time with less cost. In the future, we can explore the different policies, so as to achieve the best results in terms of response time and total cost.

REFERENCES

1. Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg, and Ivona Brandic, Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility, Future Generation Computer Systems, Volume 25, Number 6, Pages: 599-616, ISSN: 0167-739X, Elsevier Science, Amsterdam, The Netherlands, June 2009.
2. Anthony T. Velte, Toby J. Velte, Robert Elsenpeter, "Cloud Computing: A Practical Approach", The McGraw-Hill Companies (2010), [Book].
3. R. Buyya, R. Ranjan, and R. N. Calheiros, "Modeling And Simulation Of Scalable Cloud Computing Environments And The Cloudsim Toolkit: Challenges And Opportunities," Proc. Of The 7th High Performance Computing and Simulation Conference (HPCS 09), IEEE Computer Society, June 2009.
4. "Facebook,"16/6/2009;http://www.facebook.com.
5. Bhathiya Wickremasinghe, "Cloud Analyst: A Cloud Sim based Tool for Modeling and Analysis of Large Scale Cloud Computing Environments" MEDC project report, 433-659 Distributed Computing project, CSSE department., University of Melbourne, 2009.
6. www.internetworldststs.com/facebook.htm
7. "Amazon Elastic Compute Cloud (Amazon EC2)," 19/06/2009;http://aws.amazon.com/ec2/.
8. N. S. Raghava and Deepti Singh, "Comparative Study on Load Balancing Techniques in Cloud Computing," accepted in the Open journal of mobile computing and cloud computing and it is in press.